EDM Workbench User manual

Version 3.00

Prepared by:
Bon Teogene Balonzo
Gamaliel Dela Cruz
Francis Bautista

Table of Contents

Revision History 4

Introduction 5

a. Purpose 5

b. Project Scope 5

c. Definition of Terms 5

d. Overall Use Cases 7

System Overview 8

1. Title Bar 8
2. Menu Bar 8
3. Tool Bar 10
4. Data Grid 11
5. Status Bar 11

How to use the System 4

1. Import 12
2. Clipping 14

Size as Clip Type 15

Custom Sort Button 16

Time as Clip Type 16

Cancel Button 16

Save Button 17

Load Button 17

Submit Button 18

3. Sampling 18

Random Sampling 18

Stratified Sampling 19

Save Button 20

Load Button 20

Submit Button 20

4. Function Addition Process 21

Add Feature Button 21

Add Clipping 21

Add Sampling 21

Cancel Button 21

Save Button 21

Load Button 22

Run Process Button 22

5. Labeling 24

Without a Template 24

With Template 25

6. Save 27
7. Load EDM File 27
8. Export Current Tab 28

Revision History

Name	Date	Reason For Changes	Version
John Paul Contillo	20111121	First draft	V1.00
Alipio Gabriel	20111122	Edit the context of the draft	V1.00
Alipio Gabriel	20111123	Add and edit the content	V1.00
J.Contillo	20120221	User manual for version 2	V2.00
Gamaliel dela Cruz	20120526	Edit content	V3.00
Francis Bautista	20120607	Formatting and editing	V3.00

Introduction

a. Purpose

The purpose of this document is to support the user in understanding the EDM workbench's features, functions, and their usage.

This document is intended for the project stakeholders and the development team.

b. Project Scope

The EDM Workbench is a tool that helps researchers with processing data from various sources for developing meta-cognitive and behavioural models. The concept diagram in Figure 1 illustrates the system functionalities and entities interacting with it.

The EDM Workbench's functions allow users to:

- Define and modify behaviour categories of interest
- Label previously collected educational log data with the categories of interest considerably faster than current methods
- Collaborate with others in labelling data by providing ways to communicate and document labelling guidelines and standards
- Validate inter-rater reliability between multiple labellers of the same educational log data corpus
- Analyze textual data in collaborative learning situations by integrating a text categorization tool
- Automatically distil additional information from log files for use in machine learning
- Export student behaviour data to tools which enable sophisticated secondary analysis

c. Definition of Terms

Batch

A group of log files. The criteria for grouping are determined by the user.

Examples of the criteria for grouping include source and timing.

Clip

A subset of logs from a given batch.

Column

A single attribute within the dataset.

Dataset

The data from the imported files.

EDM

Educational Data Mining

Log

A record of a single action

.

Log File
A file that contains a collection of logs.
Model
A detector of meta-cognitive and motivational behaviour
Row
A set of attributes in the dataset. Usually refers to 1 log.
Interface
Refers to the system graphical user interface

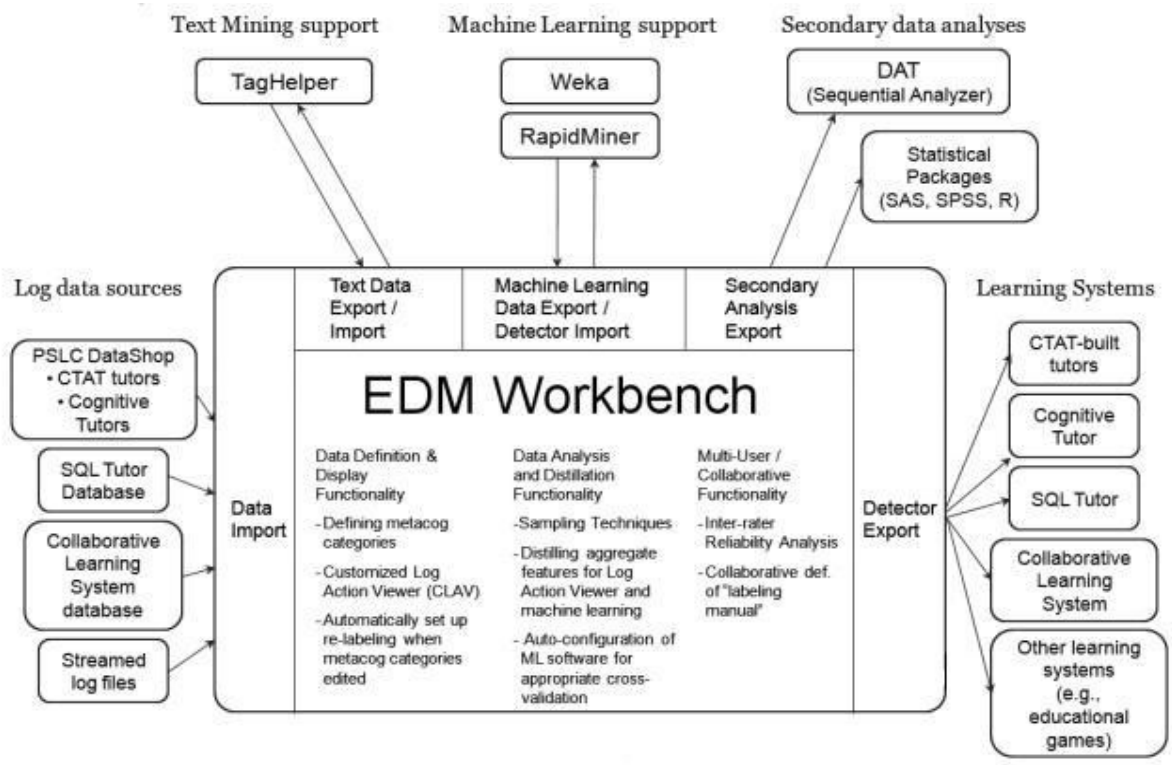


Figure 1: EDM Workbench this is

d. Overall Use Cases

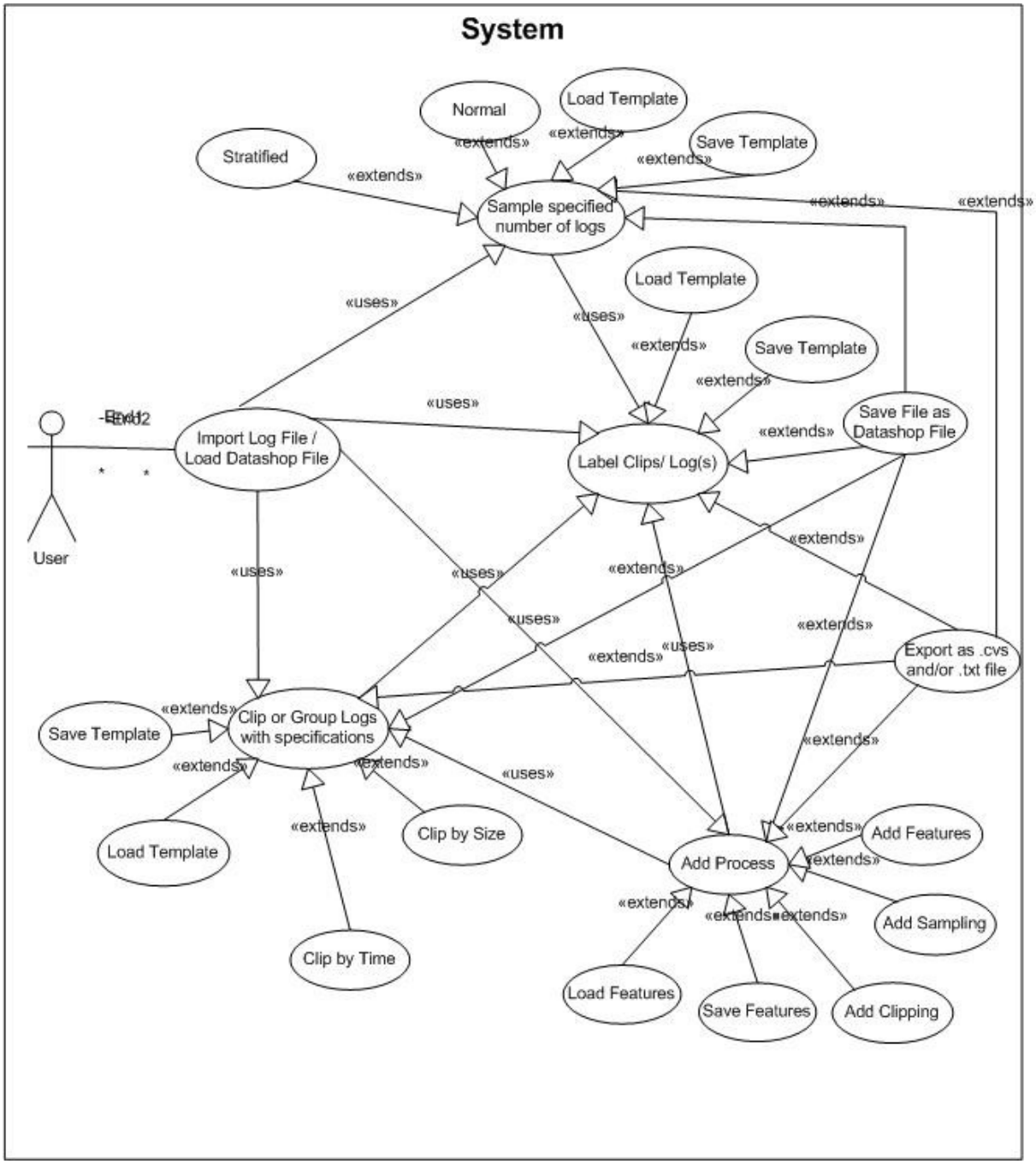


Figure 2: EDM System Map

I. System Overview

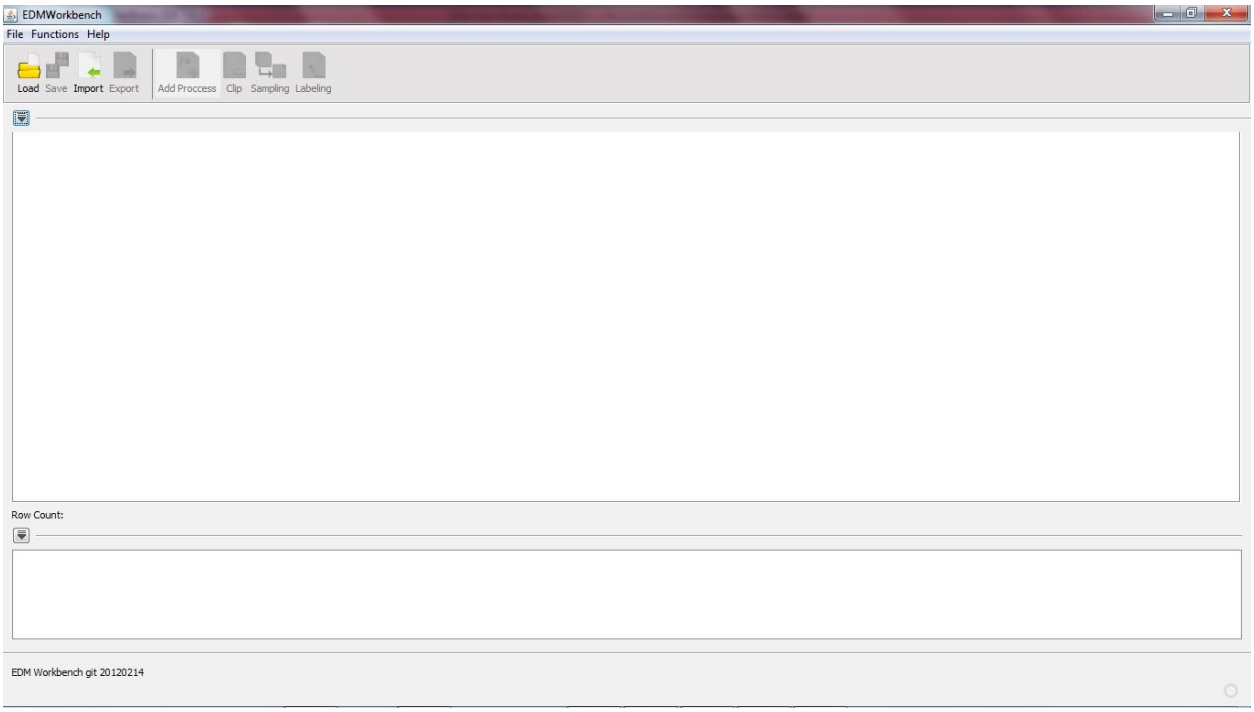


Figure 3: The EDM Workbench upon system launch

This section, discusses the interface of the system (from Top to Bottom) including its features, buttons, and functions. The discussion will only consist of a feature and function description and does not go into how these functions are carried out.

1. Title Bar

Figure 4: System Title Bar

The name of the system (may change in later versions e.g EDM Workbench version 4) is displayed here.

2. Menu Bar

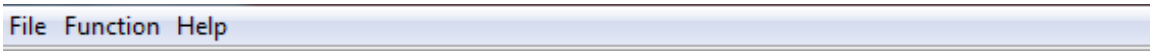
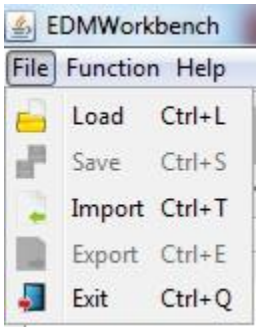
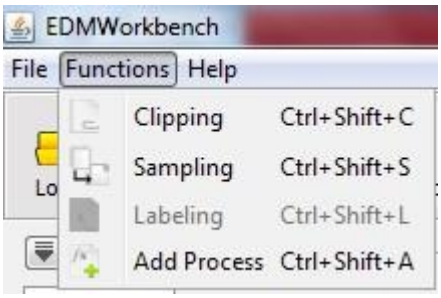


Figure 5: EDM Menu Bar

Composed of 3 Menu options (File, Function, and Help) consisting of actions buttons.



The **File Menu** is composed of 5 actions (Load, Save, Import, Export and Exit) that handle the files and logs to be displayed and/ or saved in the Data Grid.



The **Function Menu** consists of 4 log processing actions that will either be enabled or disabled depending on the state of the system.

The **Help Menu** contains the "About" action that displays the details of the system (e.g current version, system description etc).

3. Tool Bar

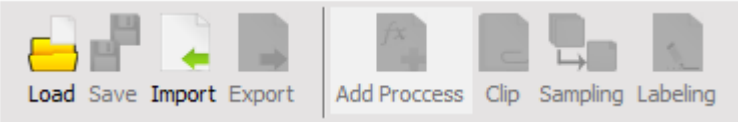


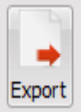
Figure 6: Toolbar

The **Tool bar** is composed of action buttons that are also found in the menu bar for ease of use.

3.1 Load Button loads logs which are in a .zip format.

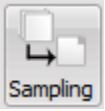
3.2 Save Button saves the logs from the active tab in the DataGrid into EDM format.

3.8 Add Features Button allows the user to add new Features for distilation of logs



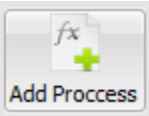
3.4 Export Button exports the final output from the active tab in the DataGrid into .cvs file formats or other specified file formats.

3.5 Clip Button groups or compiles logs from a given batch with user selected parameters.



3.6 Sampling Button creates subsets from the dataset using automatic select and grouping options

3.7 Labelling Button allows the user to label each clip.



3.8 Add Process allows the user to add multiple processes.

4. DataGrid

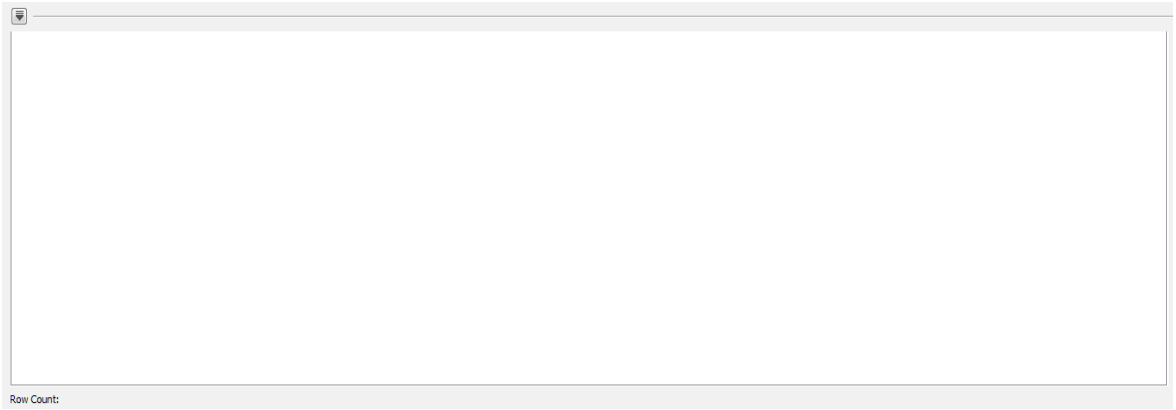
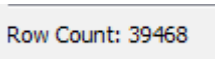


Figure 7: Data Grid

The **Data Grid** displays the logs that are active and are to be processed. The downwards-arrow button hides the data grid.



Row Count controls the amount of rows shown in the active tab

a. Status Bar

Row Count: 48443
Thu Nov 17 16:37:31 GMT-08:00 2011 - WARNING: [Feature] Incomplete input columns: features.pKnowDirect
Thu Nov 17 16:37:40 GMT-08:00 2011 - INFO: Distilled All_Data_1_export.txt
Thu Nov 17 16:38:02 GMT-08:00 2011 - INFO: Clipped All_Data_1_export.txt-clipped
Thu Nov 17 16:38:31 GMT-08:00 2011 - WARNING: [Feature] Incomplete input columns: features.FirstAttempt
Thu Nov 17 16:38:32 GMT-08:00 2011 - WARNING: [Feature] Incomplete input columns: features.pKnowDirect
Thu Nov 17 16:38:40 GMT-08:00 2011 - INFO: Distilled All_Data_1_export.txt
EDM Workbench git 20110916

Figure 8: Status Bar

The **Status Bar** displays feedback information such as status, error messages, time elapsed and others.

II. How to use the System

1. Import

Import log file by clicking Import Button

located either in File menu (Figure 5) or Toolbar (Figure 6). The system will then pop-up a dialog box asking what type of logs you want to import (CSV or Datashop file Figure 9). Click the Select Button after selecting the type of Log.

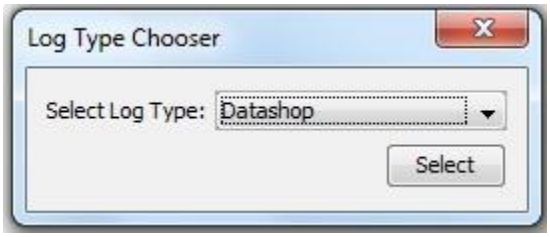
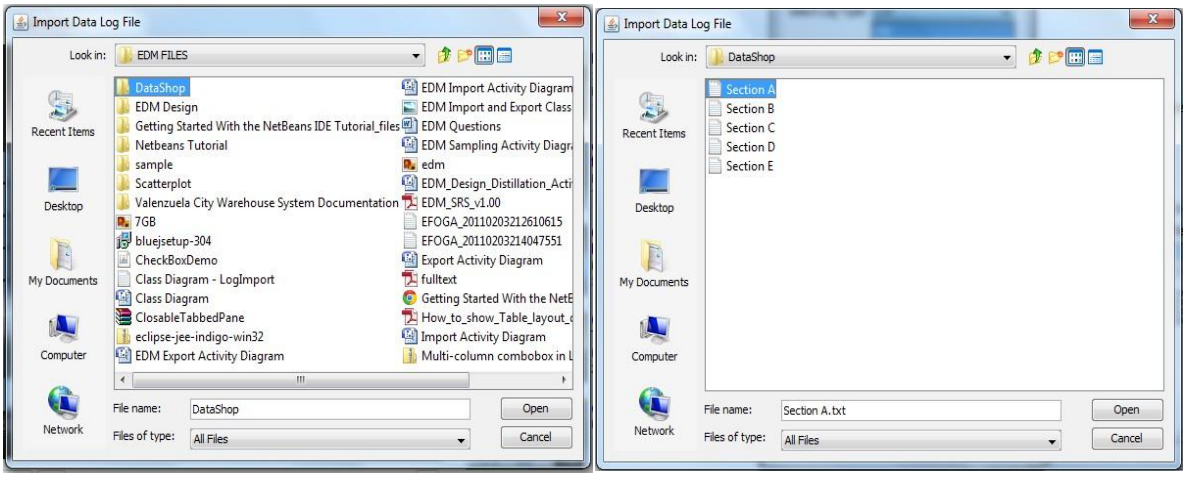


Figure 9: Log Type

Another dialog box will appear asking for the location of the log file. User may import a log file or batches of log files.



Case 1: Importing log file

After locating and choosing the log file, it will be displayed to the DataGrid (Figure 7).

Case 2: Importing batches of log files

After locating and choosing the batch of log files another dialog box will appear asking for a label describing the log files imported (e.g Logs of Students in Section A to E) (Figure 10). Click Submit Button to display the logs in the Data Grid.

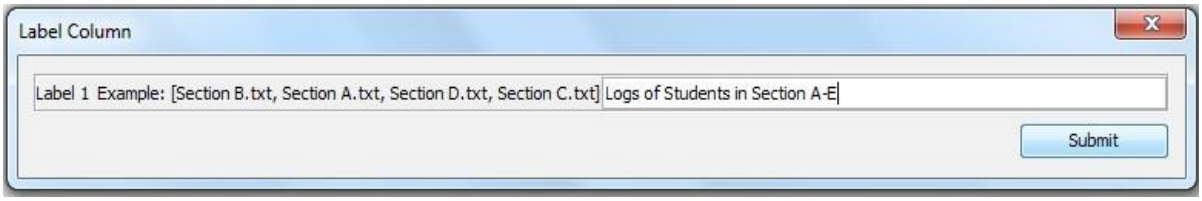


Figure 10: Label Box

The system will take some time to display the dataset depending on the log file size. The system display should be similar to that of Figure 11. All actions buttons, save for the Labelling button, should be enabled at this point.

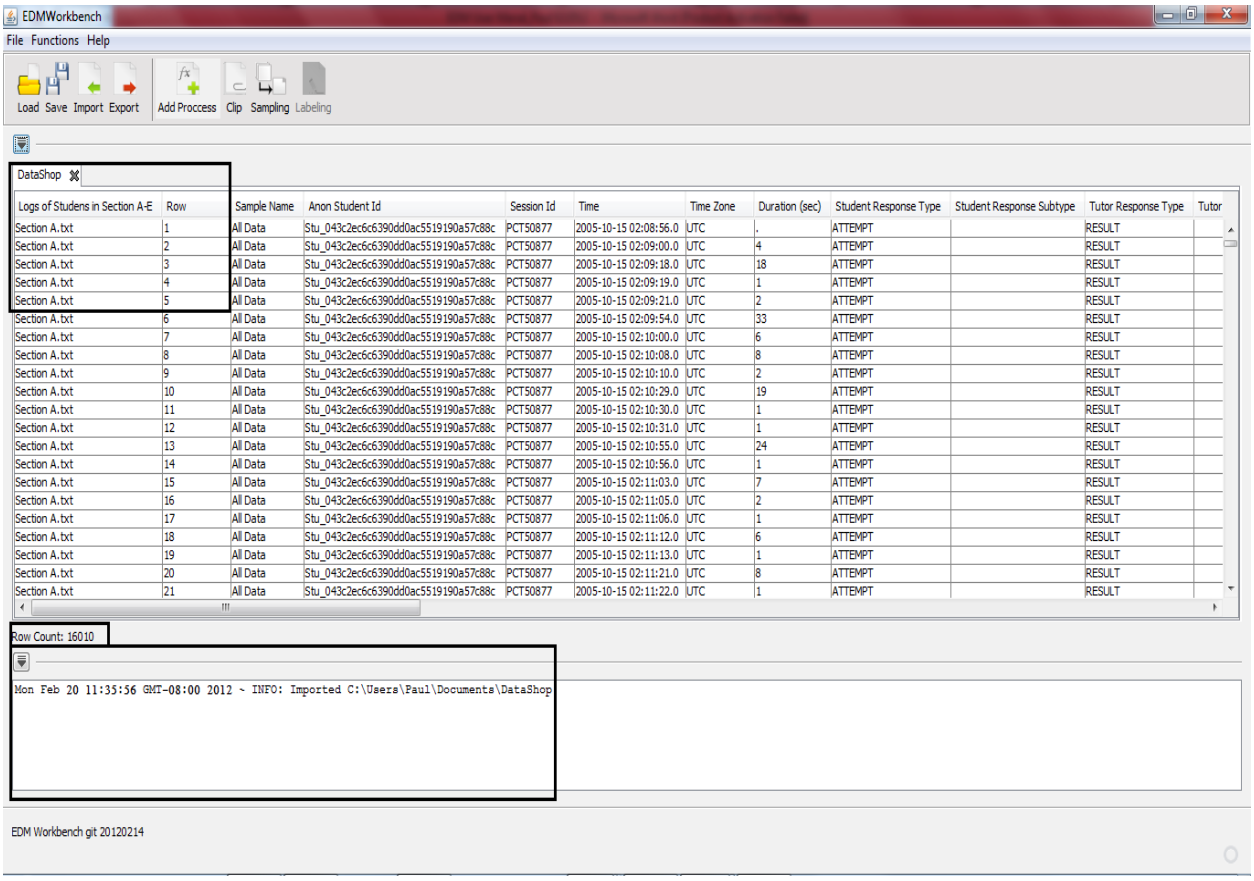

EDMWorkbench
File Functions Help
Load Save Import Export
Add Proccess Clip Sampling Labeling
DataShop
Logs of Studens in Section A-E | Row | Sample Name | Anon Student Id | Session Id | Time | Time Zone | Duration (sec) | Student Response Type | Student Response Subtype | Tutor Response Type | Tutor
Section A.txt | 1 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:08:56.0 | UTC | . | ATTEMPT | | RESULT
Section A.txt | 2 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:09:00.0 | UTC | 4 | ATTEMPT | | RESULT
Section A.txt | 3 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:09:18.0 | UTC | 18 | ATTEMPT | | RESULT
Section A.txt | 4 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:09:19.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 5 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:09:21.0 | UTC | 2 | ATTEMPT | | RESULT
Section A.txt | 6 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:09:54.0 | UTC | 33 | ATTEMPT | | RESULT
Section A.txt | 7 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:00.0 | UTC | 6 | ATTEMPT | | RESULT
Section A.txt | 8 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:08.0 | UTC | 8 | ATTEMPT | | RESULT
Section A.txt | 9 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:10.0 | UTC | 2 | ATTEMPT | | RESULT
Section A.txt | 10 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:29.0 | UTC | 19 | ATTEMPT | | RESULT
Section A.txt | 11 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:30.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 12 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:31.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 13 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:55.0 | UTC | 24 | ATTEMPT | | RESULT
Section A.txt | 14 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:10:56.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 15 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:03.0 | UTC | 7 | ATTEMPT | | RESULT
Section A.txt | 16 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:05.0 | UTC | 2 | ATTEMPT | | RESULT
Section A.txt | 17 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:06.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 18 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:12.0 | UTC | 6 | ATTEMPT | | RESULT
Section A.txt | 19 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:13.0 | UTC | 1 | ATTEMPT | | RESULT
Section A.txt | 20 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:21.0 | UTC | 8 | ATTEMPT | | RESULT
Section A.txt | 21 | All Data | Stu_043c2ec6c6390dd0ac5519190a57c88c | PCT50877 | 2005-10-15 02:11:22.0 | UTC | 1 | ATTEMPT | | RESULT
Row Count: 16010
Mon Feb 20 11:35:56 GMT-08:00 2012 ~ INFO: Imported C:\Users\Paul\Documents\DataShop
EDM Workbench git 20120214

Figure 11: Data Set in EDM UI

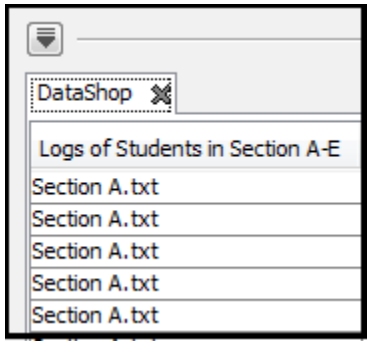


Figure 12: EDM Workbench Data Grid Tab

The "DataShop" string beside the close button is the folder name that the user specified. The label selected in the label selection dialogue (Fig.10) can be seen under the tab name.

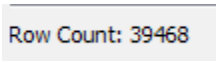


Figure 13: EDM Workbench Row Count

Row Count shows that there are 39468 records (number of rows) therein the current active tab.

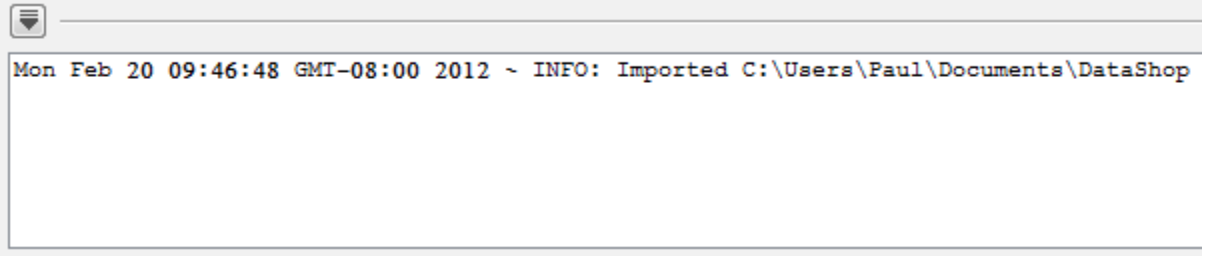


Figure 14: Status Bar with timestamp and directory

The **Status bar** displayed the information of a file imported together with the location **C:\User\Paul\Documents\Datashop** and the current time **Monday February 20 9:46 AM and 48 seconds**.

2. Clipping

To clip the dataset, click Clip Button

located either in the Function menu or Toolbar (Figure 6). When clicked, the system will display a form with the column names (the basis for grouping e.g group data with the same Logs of Student in Section A-E with the same Anon Student Id and with the same Time and so on). Clips can be divided by **Size** or **Time**.

2.1 Size as Clip

Type

By choosing **Size** as the **Clip Type**, the user will need to specify the clip size (maximum number of logs per clip). The system will then provide custom sorting options for the selected logs. When the "*complete clips only*" checkbox is checked, the system will display only clips where the number of logs is equal to the inputted clip size.

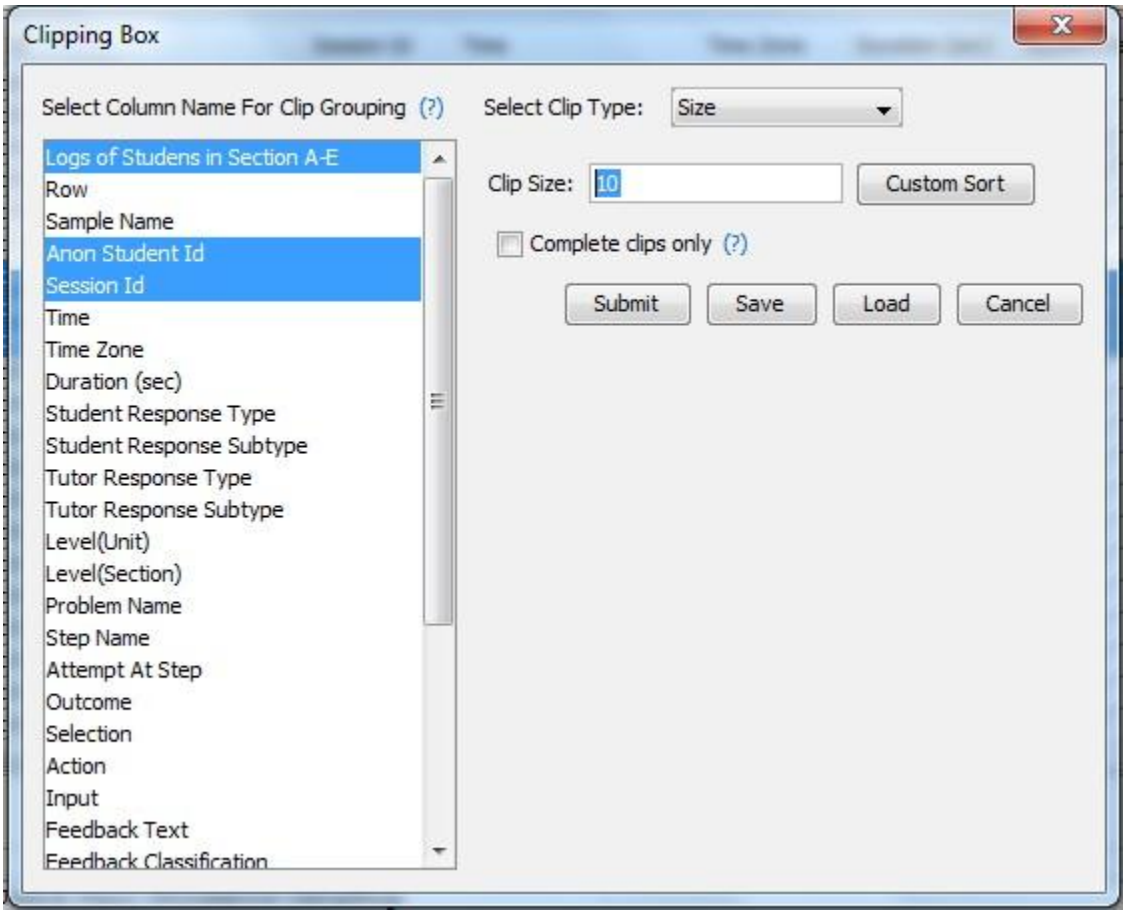


Figure 15: EDM Workbench Clipping Box

2.1.1 Custom Sort Button

This allows the user to set how the logs are formatted by sorting them. **Add Level** Button adds another row of level sorting properties and **Delete Level** deletes the selected Row. Clicking the **Submit** button will implement the selected formatting properties.

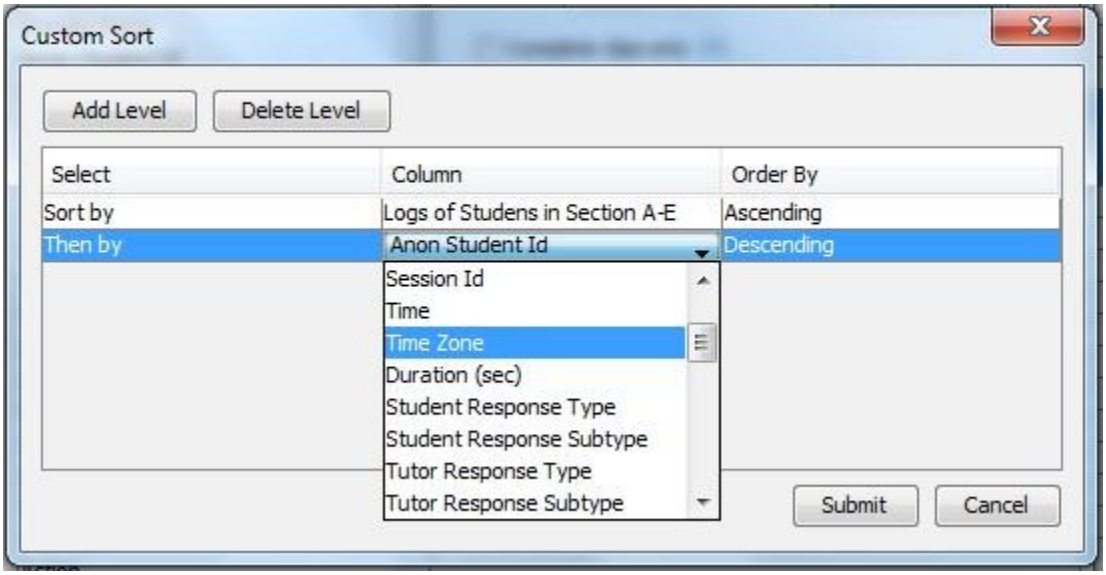


Figure 16: EDM Custom Sort

2.2 Time as Clip Type

By choosing **Time** as the **Clip Type**, the user will specify a time period per clip (e.g. 1 clip = 5 minute interval). Column name must be specified in order for clipping to work.When done, hit the submit button and the clips will be displayed on the new data grid. Double click the data grid to view the logs in the selected clip.

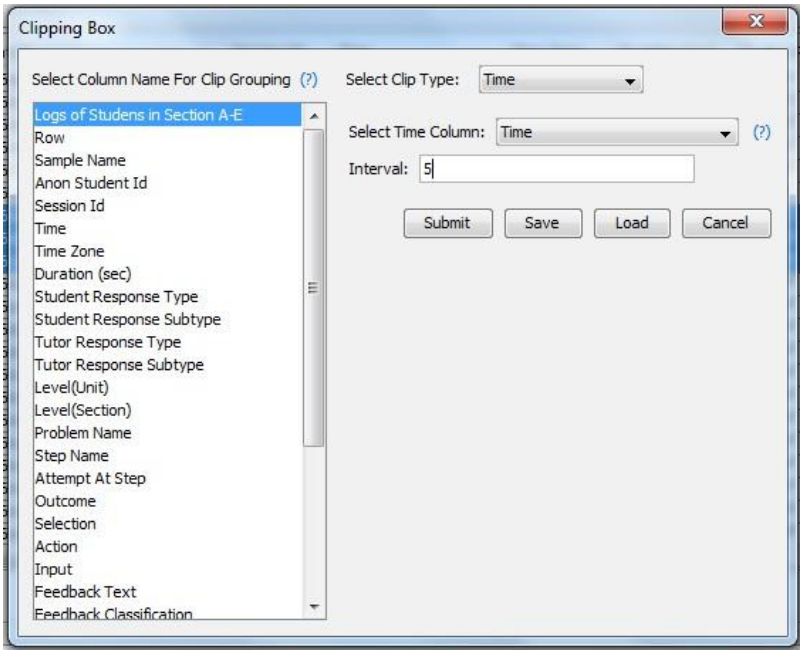


Figure 17: Time as Clip Type

2.3 Cancel Button

This cancels clipping.

2.4 Save Button

The **Save** button saves the set properties in the Clipping Form as an XML file. It will display a dialogue which will ask for a file name (see Figure 19) and when submitted, it will store the XML file in a folder which contains the XML clipping files.

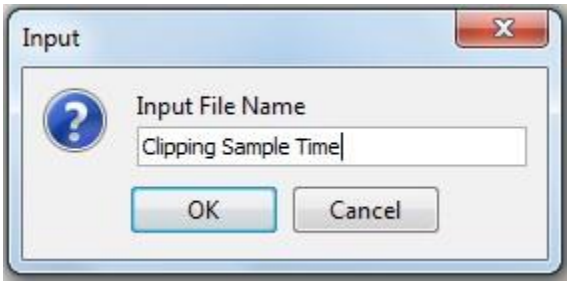


Figure 18: Save Dialogue

2.5 Load Button

Allows the user to select and choose a clipping.xml file from a list. When selected, the system loads the properties from the xml file to the system (see Figure 20 & 21).

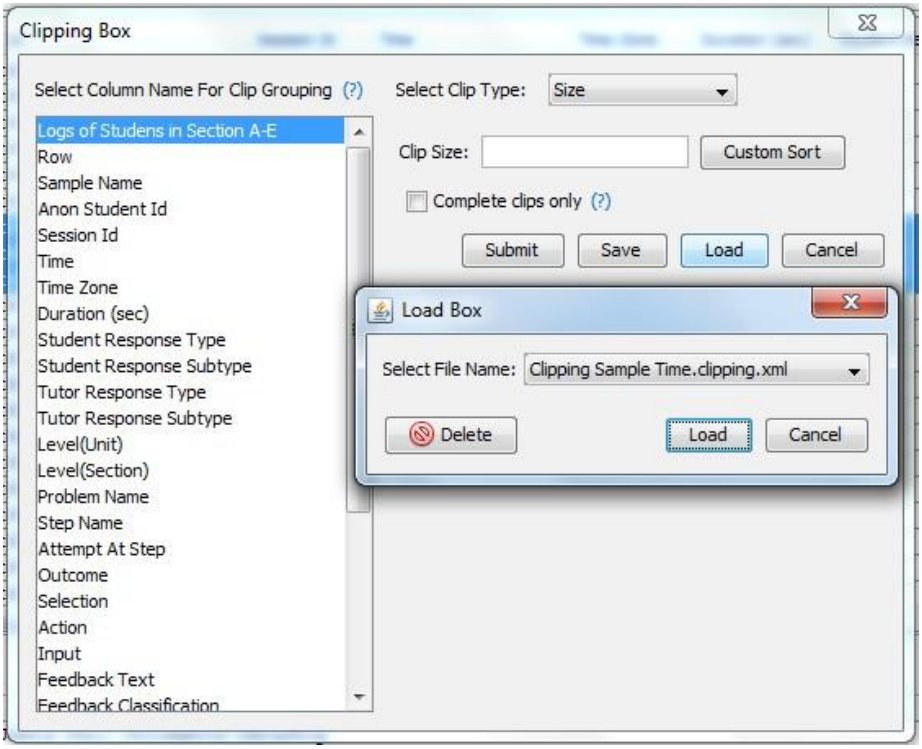


Figure 19: Initial load window

Note: From the list of clipping.xml files, the selected template is Clipping Sample Time.clipping.xml

2.6 Submit Button

This closes the Clipping Form, clips the dataset from the current tab, and displays it with its properties set in a new tab). Double click a row to view the logs within it.

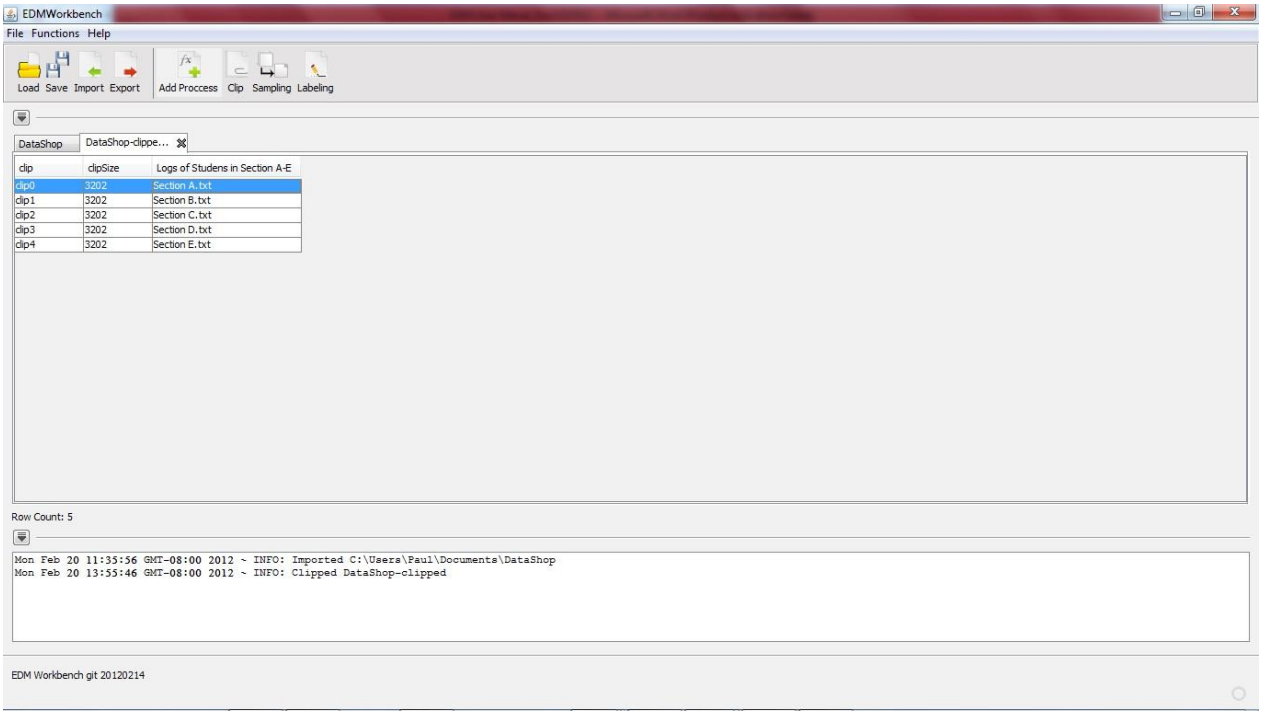

EDMWorkbench
File Functions Help
Load Save Import Export
Add Proccess Clip Sampling Labeling
DataShop
DataShop-clippe...
clip clipSize Logs of Studens in Section A-E
clip0 3202 Section A.txt
clip1 3202 Section B.txt
clip2 3202 Section C.txt
clip3 3202 Section D.txt
clip4 3202 Section E.txt
Row Count: 5
Mon Feb 20 11:35:56 GMT-08:00 2012 ~ INFO: Imported C:\Users\Paul\Documents\DataShop
Mon Feb 20 13:55:46 GMT-08:00 2012 ~ INFO: Clipped DataShop-clipped
EDM Workbench git 20120214

Figure 20: Submitting the clips

3. Sampling

To start sampling the dataset, click Sampling Button located either in the **Function** menu (Figure 5.2) or **Toolbar** (Figure 6). Sampling functionalities involve creating subsets from the dataset using automatic select and grouping options. A user may take samples or a subset from the loaded dataset and save as a new dataset. Sampling can be stratified or random.

3.1Random Sampling

In the random sampling method, the system selects samples randomly. To execute this in the Sampling interface, select the "Sampling Method" dropdown menu and click on "Normal" and set the number of samples in the Samples textbox.

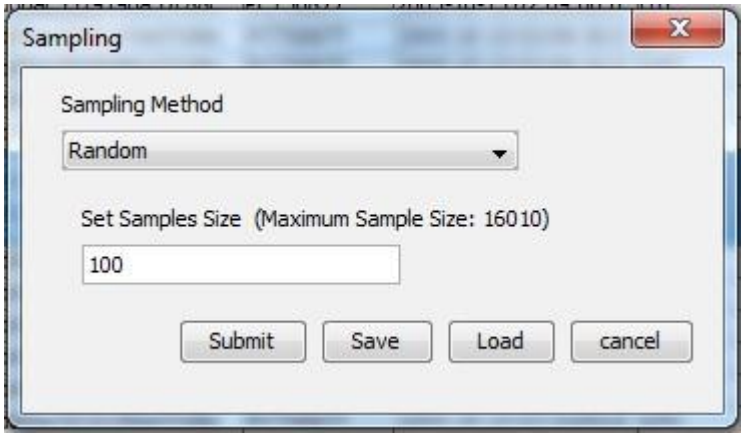


Figure 21: Sampling Method prompt

Note: The size inputted in the textbox should not exceed in the indicated maximum sample size

3.2 Stratified Sampling

For stratified sampling, users have to specify the features that define the strata from which the data will be selected. Click on the "Sampling Method" dropdown menu and select "Stratified"; set the number of samples in the Samples textbox and in the **Strata** list select the column name and log grouping (Fig. 22).

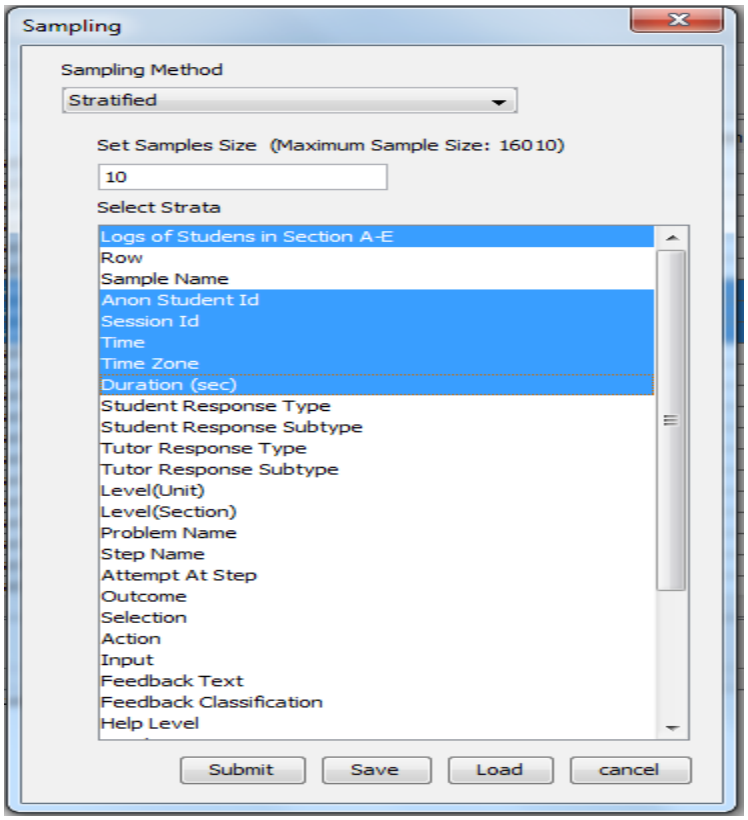


Figure 22: Stratified Sampling

3.3 Save Button

Save Button allows the user to save the set properties into a sampling.xml file.

3.4 Load Button

Through the Load button, the system allows the user to choose a template from a list and then inherit its properties.

Figure 23: Load prompt

3.5 Submit Button

It closes the Sampling Form, implements the sampling process and then displays the result in a new tab.

4. Add Process

This allows the user to add multiple processes and run them in sequence.

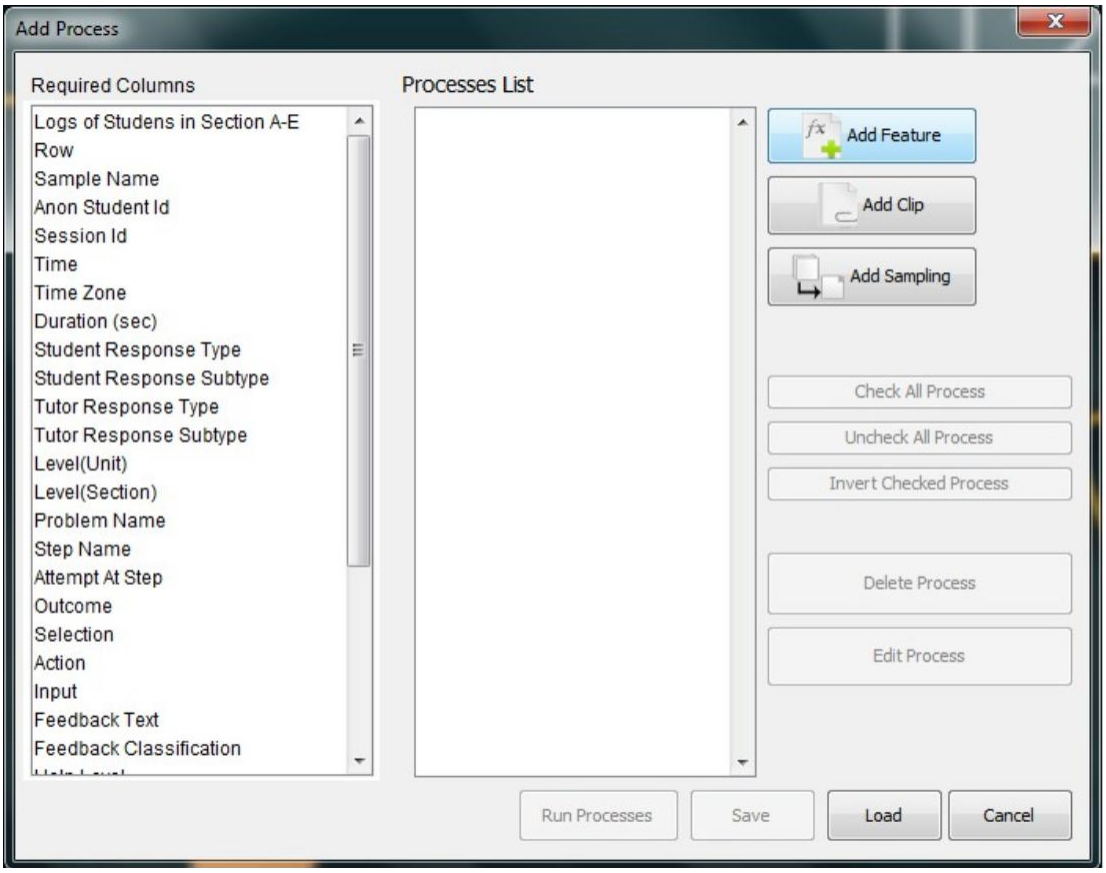


Figure 24: Add Feature Screen

4.1 Add Feature Button

Allows user to add a function and edit its functions.

4.2 Add Clipping

Allows user to set the desired clipping properties (See Clipping) and when done, the form applies the selected properties in the clipping form.

4.3 Add Sampling

Allows user to set desired sampling properties (See Sampling) and when done, the form gets the sampling properties set in the sampling form.

4.4 Cancel button

Cancels and closes the Add Process form.

4.5 Save Button

The system shall save all the properties set in the Processes List which are checked in a process.xml file.

4.6 Load Button

The system will load the user selected process.xml file upon clicking the load button.

4.7 Run Process Button

The system runs all checked processes (distil clipping and/or sampling) in the process list. The system shall display information feedback in the Status Bar on what process it is currently taking and throws an error dialogue when the system encounters error.

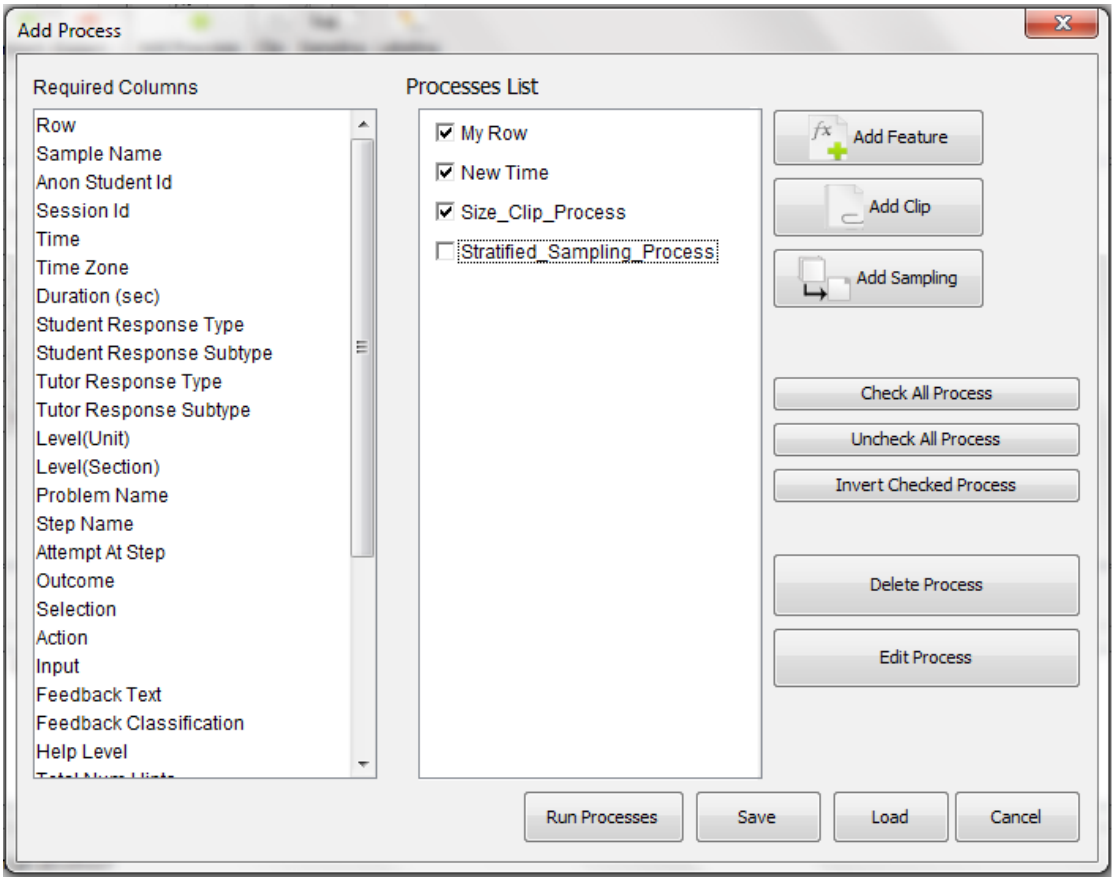
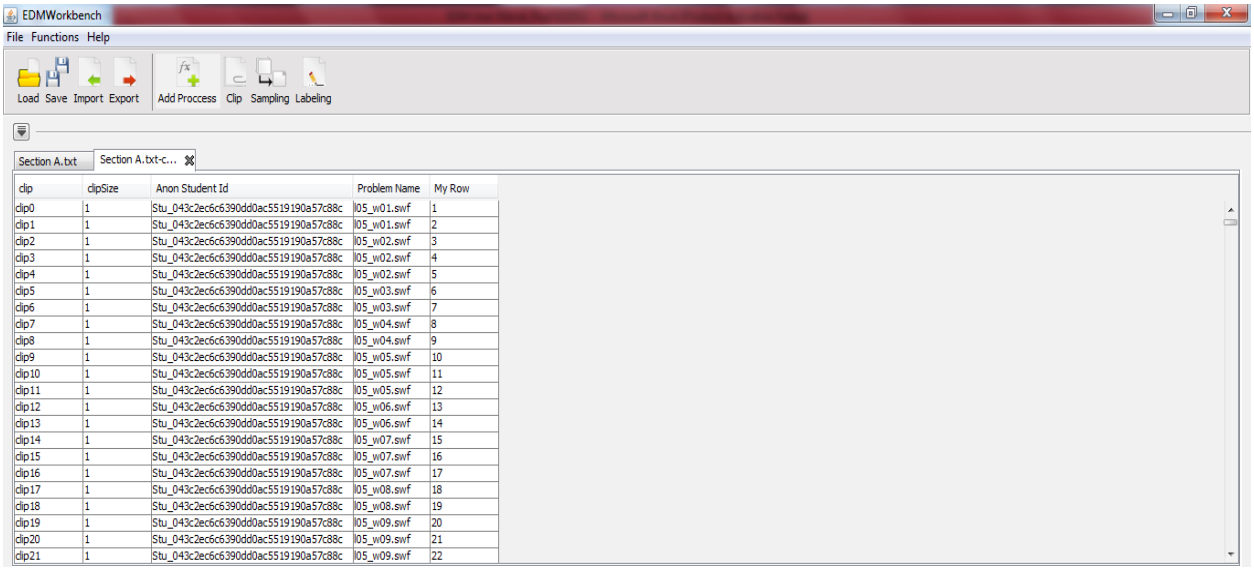


Figure 25: Process list



clip	clipSize	Anon Student Id	Problem Name	My Row
clip0	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w01.swf	1
clip1	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w01.swf	2
clip2	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w02.swf	3
clip3	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w02.swf	4
clip4	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w02.swf	5
clip5	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w03.swf	6
clip6	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w03.swf	7
clip7	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w04.swf	8
clip8	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w04.swf	9
clip9	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w05.swf	10
clip10	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w05.swf	11
clip11	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w05.swf	12
clip12	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w06.swf	13
clip13	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w06.swf	14
clip14	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w07.swf	15
clip15	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w07.swf	16
clip16	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w07.swf	17
clip17	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w08.swf	18
clip18	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w08.swf	19
clip19	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w09.swf	20
clip20	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w09.swf	21
clip21	1	Stu_043c2ec6c6390dd0ac5519190a57c88c	l05_w09.swf	22

Figure 26: Clipping display

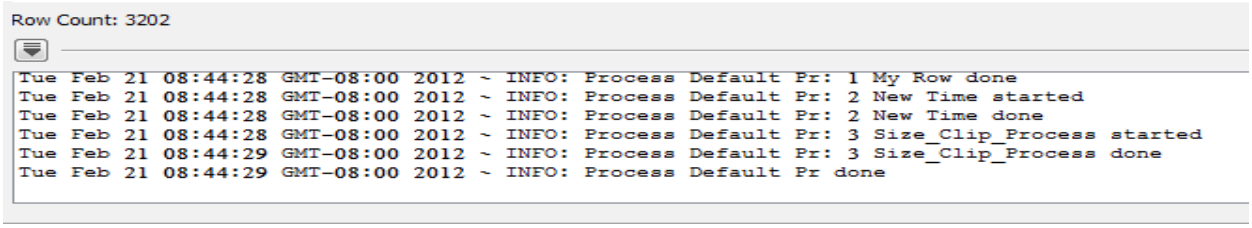

Row Count: 3202
Tue Feb 21 08:44:28 GMT-08:00 2012 ~ INFO: Process Default Pr: 1 My Row done
Tue Feb 21 08:44:28 GMT-08:00 2012 ~ INFO: Process Default Pr: 2 New Time started
Tue Feb 21 08:44:28 GMT-08:00 2012 ~ INFO: Process Default Pr: 2 New Time done
Tue Feb 21 08:44:28 GMT-08:00 2012 ~ INFO: Process Default Pr: 3 Size_Clip_Process started
Tue Feb 21 08:44:29 GMT-08:00 2012 ~ INFO: Process Default Pr: 3 Size_Clip_Process done
Tue Feb 21 08:44:29 GMT-08:00 2012 ~ INFO: Process Default Pr done

Figure 27: Clipping run feedback

KC(Unique...	KC Catego...	School	Class	My Row	New Time
KC696		CMU		1	2005-10-15 ...
KC814		CMU		2	2005-10-15 ...
KC1592		CMU		3	2005-10-15 ...
KC238		CMU		4	2005-10-15 ...
KC1422		CMU		5	2005-10-15 ...
KC1415		CMU		6	2005-10-15 ...
KC1356		CMU		7	2005-10-15 ...
KC1329		CMU		8	2005-10-15 ...
KC75		CMU		9	2005-10-15 ...
KC496		CMU		10	2005-10-15 ...
KC8		CMU		11	2005-10-15 ...
KC1410		CMU		12	2005-10-15 ...
KC1547		CMU		13	2005-10-15 ...
KC1330		CMU		14	2005-10-15 ...
KC750		CMU		15	2005-10-15 ...
KC808		CMU		16	2005-10-15 ...
KC658		CMU		17	2005-10-15 ...
KC1397		CMU		18	2005-10-15 ...
KC668		CMU		19	2005-10-15 ...
KC742		CMU		20	2005-10-15 ...
KC1143		CMU		21	2005-10-15 ...

Figure 28: Distil features

5. Labelling

Loads the Labelling Form and asks the user to specify Labelling parameters, other additional information, and the actual labels. The user may use a Template label or not. The "Use Template" checkbox is selected by default.

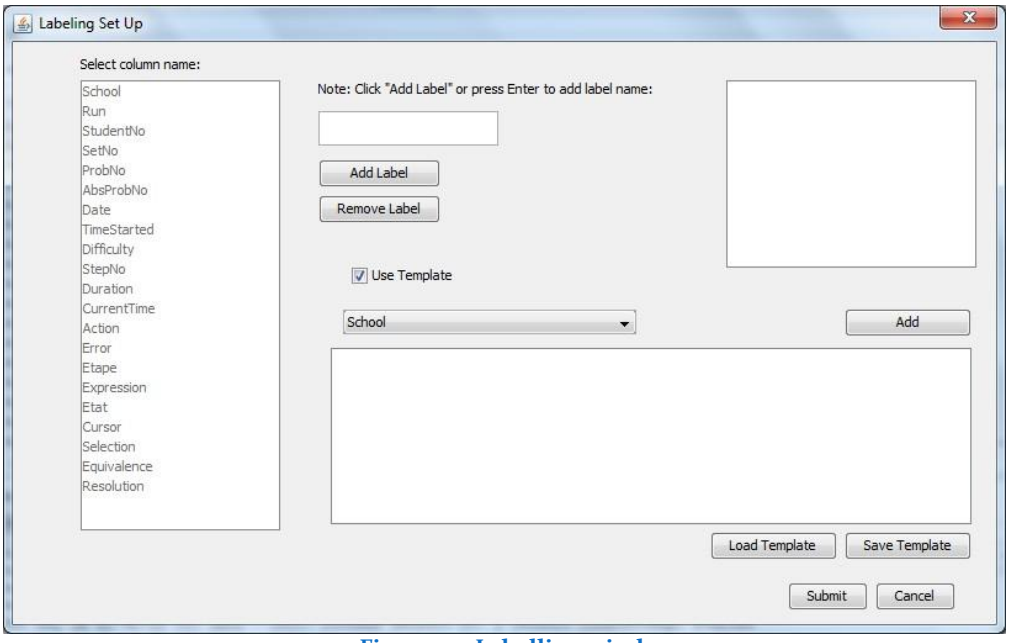


Figure 29: Labelling window

5.1 Labelling without a Template

User may not use a template by unchecking the "Use Template" checkbox. The system will require the user for input in the Labels and Required columns only.

A. Set-Up Labelling parameters

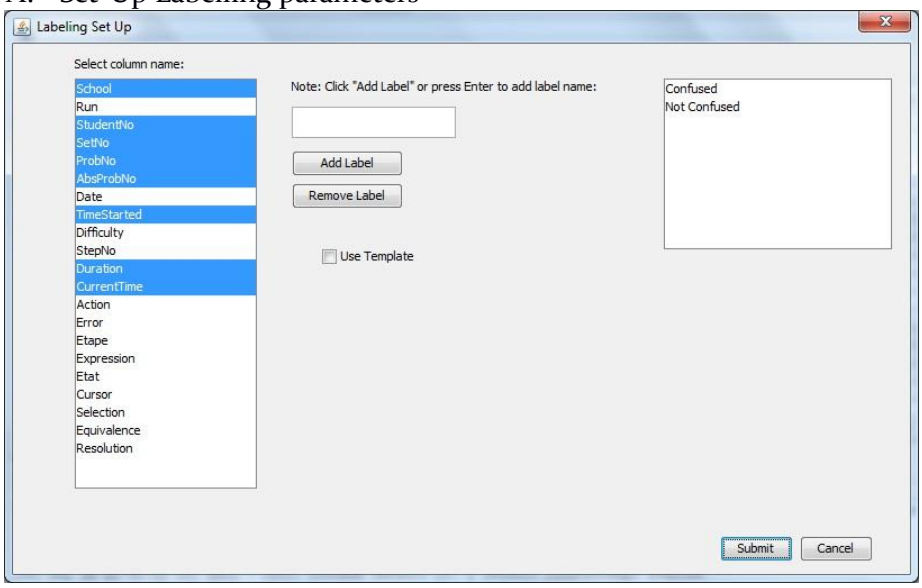


Figure 30: Labelling window w/o template

B. Labelling the dataset

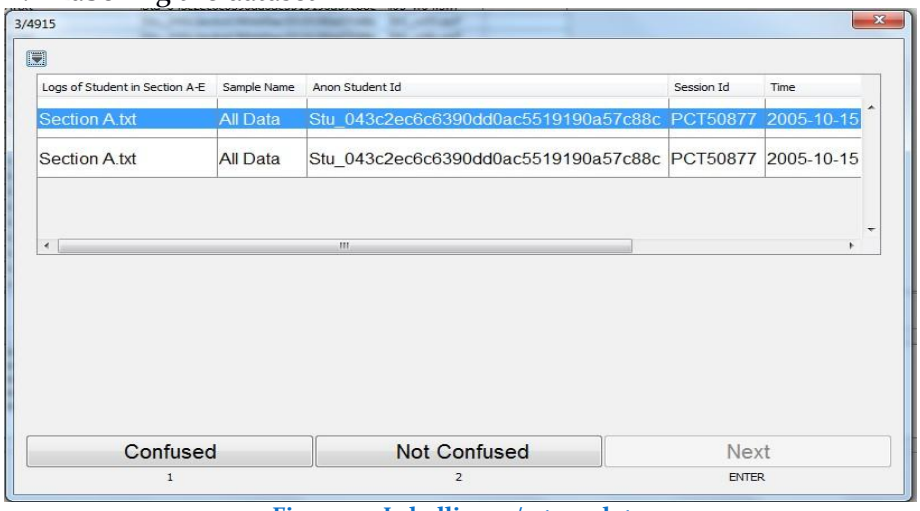


Figure 31: Labelling w/o template

5.2 Using a Template

Using template allows the user to load or manually set a desired template by selecting the required columns, typing in labels and constructing sentences.

A. **Set-up Labelling parameters**

1. ***Add Button***

In constructing sentences that describe a clip, users can manually type in the features by enclosing them in brackets "[]". Users can also select a feature from the dropdown list and click the Add button to add it to the sentence.

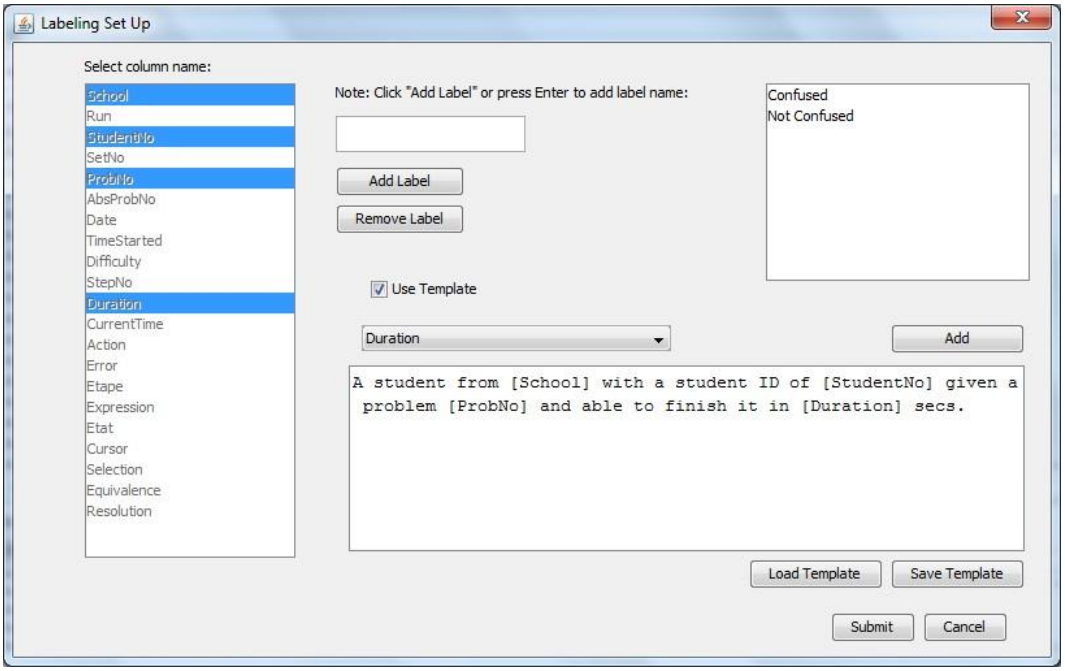


Figure 32: Parameter addition

Note: The system will automatically select the parameter in the textbox to the "Select Column Name" list.

2. ***Save Template***

   The system allows the user to save the set Labelling properties. The system will then pop-up a dialogue and ask for the desired name for the template and then save the file as a Labelling.xml file.

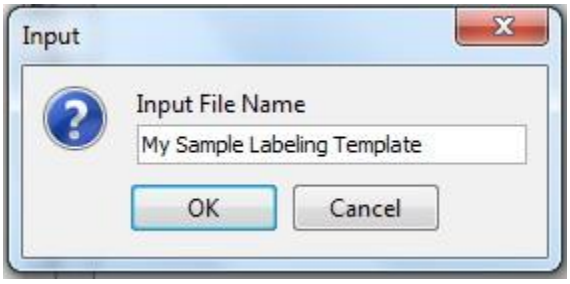


Figure 33: File Name input window

3. ***Load Template***

   The user may select a template from the list of labelling templates displayed by the system. The system will then load the properties from the selected template to the labelling form.

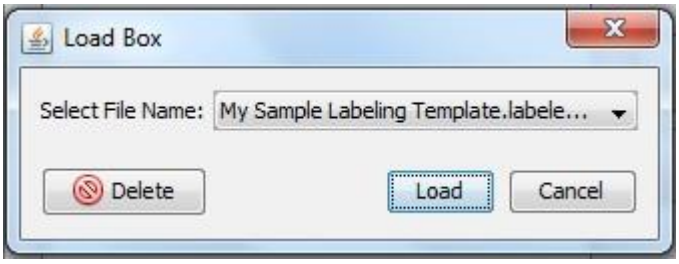


Figure 34: Label template loading

B. **Labelling the dataset**

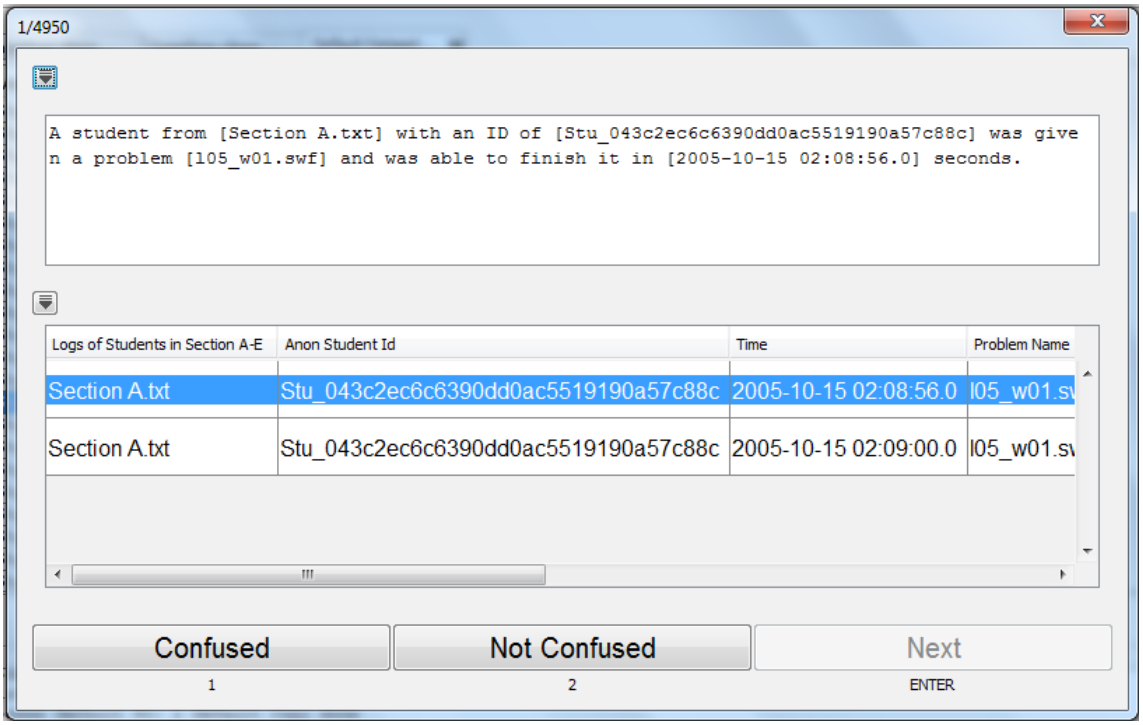


Figure 35: Dataset labelling window

6. Save

Saves the dataset in the current tab by clicking the Save button

located either in **File** menu (Figure 5.1) or **Toolbar** (Figure 6). The system will ask for the directory and then save it in zip format.

Note: Saving files will take time depending on the size of the dataset and speed of the computer.

7. Load EDM File

Loads EDM files by clicking the load button

located either in the **File** menu (Figure 5 .1) or **Toolbar** (Figure 6). Error dialogues will be displayed if any error is found with the specified directory or file.

Note: The action button will be enabled depending on the file loaded.

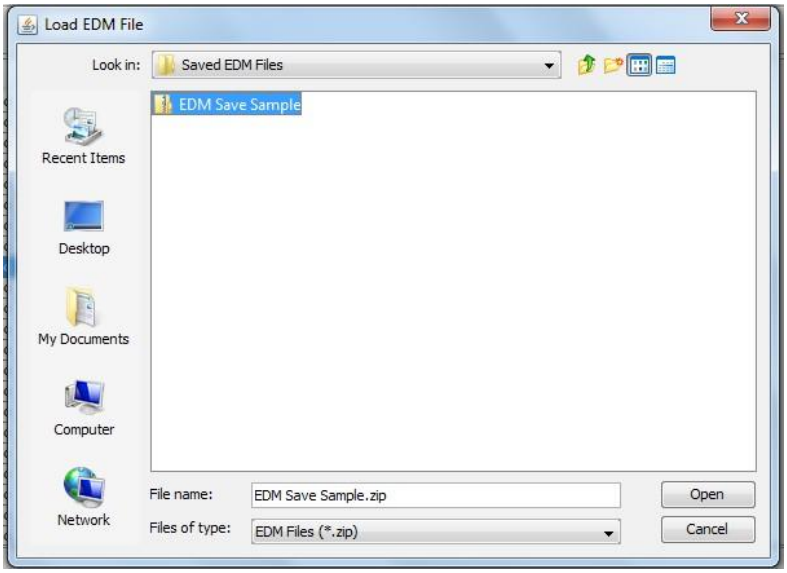


Figure 36: EDM file-load window

8. Export Current Tab

By clicking the export button

located either in the **File** menu (Figure 5.1) or **Toolbar** (Figure 6), the system will save the current active tab into a .csv file or into another specified format. Users must specify the directory in which the file will be saved.

Note: Exporting a file will take time depending on the dataset's size.

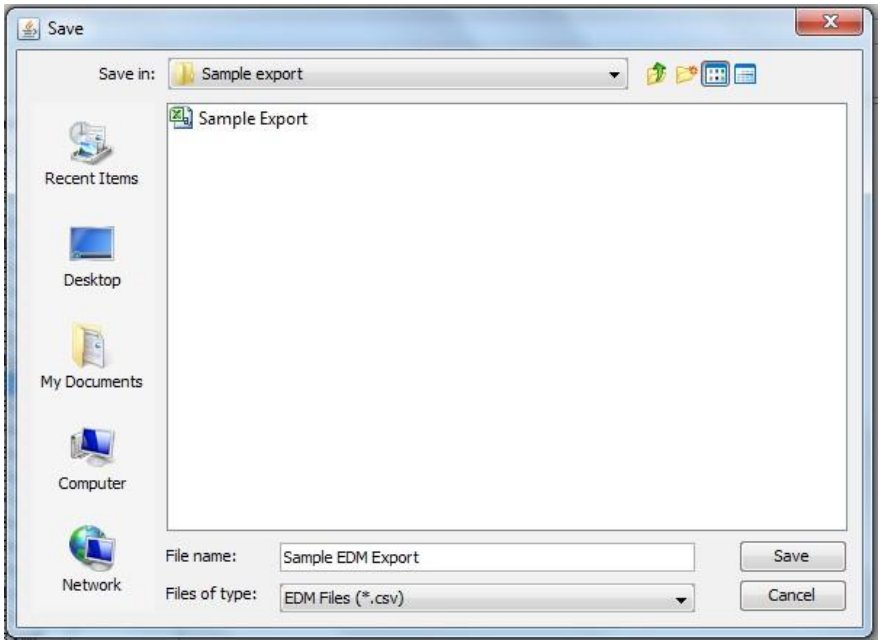


Figure 37: Export window